ELSEVIER

Molecular and Biochemical Parasitology 90 (1997) 69–79

MOLECULAR AND BIOCHEMICAL PARASITOLOGY

Genetic and biochemical analysis of erythrocyte-stage surface antigens belonging to a family of highly conserved proteins of *Babesia equi* and *Theileria* species

D.P. Knowles [a,b,*], L.S. Kappmeyer [a], L.E. Perryman [c]

[a] *Animal Disease Research Unit, Agricultural Research Service, US Department of Agriculture, Pullman, WA 99164-7030, USA*
[b] *Department of Veterinary Microbiology and Pathology, Washington State University, Pullman, WA 99164-7040, USA*
[c] *Department of Microbiology, Pathology, and Parasitology, North Carolina State University, Raleigh, NC 27606, USA*

Received 2 July 1997; accepted 8 August 1997

Abstract

Erythrocyte-stage *Babesia equi* expresses a 34-kDa immunodominant antigen recognized by antibody from persistently infected horses worldwide. This erythrocyte-stage surface protein, equi merozoite antigen-1 (EMA-1) is encoded by a single copy gene, and was previously shown to share 33% amino acid identity with similar sized proteins of *Theileria sergenti* and *T. buffeli*. A mean homology of 31% amino acid identity extends to similar sized proteins of *T. parva*, *T. annulata* and *T. mutans*. Genomic and cDNA copies of a second *B. equi* gene, *ema2* were cloned. The single copy *ema2* gene encodes a 30-kDa protein (EMA-2) that shares 52% amino acid identity with EMA-1. EMA-2 also shares a mean amino acid identity of 31% with proteins of similar molecular mass from *Theileria* species. EMA-1 and EMA-2 each contain a glycosylphosphatidylinositol anchor. These unique erythrocyte-stage surface proteins of *B. equi* and *Theileria* species lack antigenic repeats, and excluding the signal peptide, contain one or no cysteines. Consistent with the hypothesis that this family of proteins interacts with the erythrocyte surface, the *T.* species proteins possess a basic isoelectric point. The *B. equi* proteins have acidic isoelectric points, but 24-mer peptides within them have strongly basic net charges. © 1997 Elsevier Science B.V.

Keywords: Babesia equi; *Theileria*; Immunodominant surface antigens; Selection; Charge analysis; Protein structure

1. Introduction

Babesia equi, a member of the phylum Apicomplexa, is a tick-transmitted protozoan parasite of horses. Infection of susceptible horses results in

* Corresponding author. Tel.: +1 509 3356022; fax: +1 509 3358328; e-mail: dknowles@vetmed.wsu.edu

PII S0166-6851(97)00150-3

anemia, which is caused by replication of the erythrocyte-stage parasite. Erythrocyte destruction does not require specific immune responses by infected horses [1,2]. The majority of horses survive the anemia, which follows initial infection, and become persistently infected life-long carriers. Although immunocompetent horses persistently infected with *Babesia equi* are able to control parasite replication, the immune mechanisms mediating this control are not defined.

We showed previously that immunocompetent young horses infected with *Babesia equi* preferentially produce antibodies to erythrocyte-stage antigens of 30- and 34-kDa during resolution of acute infection [1,2]. The 34-kDa antigen, equi merozoite antigen-1 (EMA-1), shares 33% amino acid identity and 56% amino acid similarity with a 33-kDa protein of *Theileria sergenti* [3]. The *ema1* gene product also shares similar homology with a 34-kDa protein of *T. buffeli* [3].

Amino acid sequences for related erythrocyte-stage proteins of similar molecular mass to EMA-1 have been reported for *T. parva*, *T. annulata*, *T. sergenti*, and *T. buffeli* [4–6]. This group of erythrocyte-stage *T.* species proteins are characterized by their proposed surface location, molecular masses of 30–34 kDa, and lack of apparent antigenic repeats. These proteins contain one or no cysteines, and all induce high levels of antibody.

Infected horses worldwide respond with high levels of antibody to a surface exposed, conserved protein epitope within EMA-1 and a 30-kDa antigen, similar to the vigorous antibody response reported to homologous proteins in *Theileria* species [4,6,7]. The worldwide conservation of the EMA-1 epitope defined by monoclonal antibody 36/133.97 and the presence of protein homologs in *Theileria* species led to the hypothesis that the invariant residues of these proteins determined their structure and function [3]. Recent reports indicate that genes encoding *Theileria* protein homologs exhibit areas of marked genetic diversity [4,8–10] and that putative N-linked glycosylation sites within the *T. annulata*, *T. sergenti*, *T. buffeli* and *T. parva* protein homologs encoded by these genes are diverse [4].

In consideration of the above data, we obtained PCR copies of the *ema1* gene from a *B. equi* infected horse in Brazil and a donkey in Morroco. Analysis of these *ema1* clones and genomic and cDNA clones of a second related *B. equi* gene, encoding a 30-kDa antigen related to EMA-1, allowed us to address the following questions: (1) Does the previous genetic homology demonstrated between EMA-1 and *T. sergenti* and *T. buffeli* extend to other *T.* species?; (2) What is the extent of the genetic diversity of the gene encoding EMA-1?; (3) What is the genetic relationship between the genes encoding EMA-1 and a second immunodominant *B. equi* erythrocyte stage surface protein, EMA-2?; (4) Are the putative N-linked glycosylation sites of these *B. equi* proteins conserved among isolates and between each other?; and (5) do EMA-1 and EMA-2 possess a glycosylphosphatidylinositol anchor and N-linked glycosylation? The purpose of this report is to present data providing answers to these five questions.

2. Material and methods

2.1. Genomic and cDNA cloning of ema2 gene

The *ema2* cDNA was identified from an expression library as reported previously [11]. The expression library was screened with mAb 36/133.97. Positive signals were rescreened until a purified population of recombinant phage was obtained, then excised to recover the recombinant pBluescript plasmid (Stratagene).

B.equi Florida genomic DNA was isolated from infected erythrocytes by incubation in proteinase K and extraction with buffered phenol and phenol/chloroform/isoamyl alcohol. DNA was digested to completion with *Eco*R1 and ligated into the Lambda ZAPII cloning phage. The packaged library was screened by hybridization by the [^{32}P]dCTP labeled *ema2* gene, liberated from the plasmid by an *Eco*R1 digest and purified from low melting temperature agarose. Positive signals were repeatedly screened until a purified population of phage

was obtained. The recombinant pBluescript plasmid was derived by in vivo excision (Stratagene).

Both strands of inserts were sequenced full length using the Sequenase kit (USB).

2.2. Database searching and sequence analysis

Databases were searched by BLAST [12] through NCBI. Sequences were manipulated and alignments made using GCG programs on a UNIX computer. A profile was constructed from sequences shown in Fig. 1 (GenBank™ accession numbers L47209, S54804, S27859, S27863, U85039, L13784 and U97166) representing *T. parva*, *T. annulata*, *T. buffeli*, *T. sergenti*, *T. mutans*, EMA-1 and EMA-2, respectively. An additional *T.* species sequence from China was also used (GenBank™ accession number D50305). The profile was used in the GCG program Profile Search to look for functional relatedness to other proteins within the database.

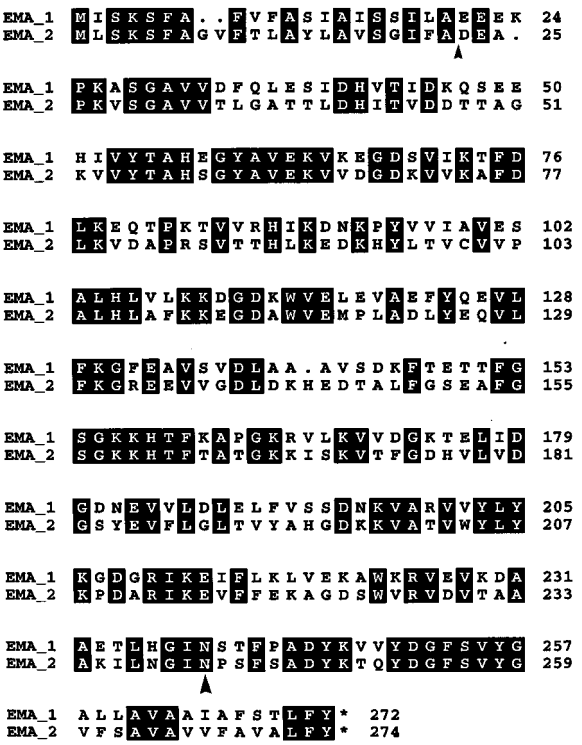

Fig. 1. Alignment of the deduced amino acid sequences for *B. equi* erythrocyte-stage proteins EMA-1 (GenBank™ accession number L13784) with EMA-2 (GenBank™ accession numbers U97166 (cDNA) and U97169 (genomic)). The GCG program PILEUP was used to align the sequences, and PRETTYBOX (public domain, R. Westerman, AIDS Center Laboratory for Computational Biochemistry, Purdue University) was used to display the alignment. Identical residues are designated with a black box. Gaps are represented by dots. Arrows indicate the signal peptidase cleavage and potential N-linked glycosylation sites, respectively.

2.3. PCR cloning of ema1 genes from infected horses in Morocco and Brazil

Genomic DNA was isolated from whole blood collected from *B. equi* carrier animals by venipuncture. One blood sample was from a naturally infected donkey located in Gharb, Morocco and the second sample from an experimentally infected horse from Porto Alegre, Brazil. This horse was infected intravenously with the Pelotas-1 isolate obtained from a naturally infected horse within Brazil. The blood was collected into EDTA and DNA isolated using a commercially available kit (Puregene, Gentra Systems, USA). *B. equi* genomic DNA for amplification of both the Morocco and Brazil *ema1* genes was isolated and purified in the respective countries. PCR primers (5′GATGATTTCCAAATCCTTTGC3′ and 3′GTAAAAGATGAGATAAAATGA5′) were chosen based on Florida *ema1* sequence (GenBank™ accession number L13784) in order to amplify full-length clones. One hundred microliter PCR reactions consisted of 4 μ1 of 50 ng μl^{-1} each primer, 1μl (approximately 1 μg) DNA, 5 μl Master mix (BM) and 45 μl water. Reactions cycled at 95°C for 5′, 30 cycles at 95°C for 30′, at 52°C for 20′, at 72°C for 30′ and then 72°C for 5′. Reactions were precipitated and resuspended in one-tenth volume. Controls for the PCR reactions included reagent controls (all components added except DNA). These controls were consistently negative for amplification. Two microliters were ligated into pCR3.1. TA cloning vector (Invitrogen) and 1 μl of ligation reaction was used to transform competent *E. coli* strain OneShot

TOP10F′ (Invitrogen). Both strands of one clone from each PCR reaction were sequenced for their full-length using automated sequencing technology (Applied Biosystems).

2.4. B. equi cultures and in vitro labeling

B. *equi* cultures were established from infected blood from a splenectomized pony when the parasitemia was at 3.0% and ascending. Normal equine erythrocytes were from an uninfected adult female Arabian horse, collected on glass beads for defibrination, and washed in Dulbecco's PBS supplemented with 15 mM EDTA. Tissue culture medium was as previously reported [13], but supplemented with 20 mM Hepes and adjusted to pH 7.2. Cultures were split 1:2 daily to maintain a peak parasitemia of approximately 5–6%, and maintained at 37°C in 2% O_2, 5% CO_2, and 93% N_2. Tissue culture parasites were labeled with various fatty acids or sugars obtained from DuPont NEN. Compounds supplied in ethanol were speed vacuumed to drive off the solvent and then resuspended in tissue culture medium ([^{3}H]fucose, [^{3}H]galactose, [^{14}C]ethanolamine), or further processed as follows ([^{3}H]myristic acid, [^{3}H]palmitic acid). Dried fatty acid pellets were resuspended in 20 μl ethanol, then added to 250 μl of 20 mg ml^{-1} defatted BSA (Sigma # A7511) in 10 mM phosphate buffer at pH 7.0 containing 150 mM NaCI. The mixture was vortexed and 3.75 ml tissue culture medium was added. [^{3}H]inositol and [^{3}H]glucosamine (aqueous solutions) were added to tissue culture medium at a rate of 1 part label to 5 parts medium.

For labeling, individual wells were split 1:2, allowed to grow for 8 h and then the medium was removed from the top of the settled cell layer. Medium (1 ml) containing radiolabeled precursors was applied to selected wells and allowed to incubate for 16 h.

2.5. Immunoprecipitation of labeled products

Labeled cultures were centrifuged 800 × *g* for 15′ at 4°C and the supernatants were discarded. Pelleted erythrocytes were washed four times using ice-cold PBS, centrifuged under the same conditions, and the final pellet was resuspended in lysis buffer [14] containing 1% NP40. After 1 h incubation on ice, the material was frozen to −80°C. After being thawed rapidly to 37°C, each sample was sonicated 15″ in a cup sonicator. Products were centrifuged 150 000 × *g* for 1 h, the supernatants recovered and stored at −20°C until further used.

Immunoprecipitations were performed as previously described [14] with each lysate being equally distributed between antibodies. Horse serum was used at the rate of 10 μl undiluted and mAbs used at a rate of 10 μg ml^{-1}. Precipitates were boiled in SDS-PAGE sample buffer and separated on 7.5–17.5% polyacrylamide gels prior to autoradiography of dried gels.

3. Results

3.1. Analysis of the B. equi gene encoding equi merozoite antigen-2 (EMA-2)

Genomic and cDNA clones of a second *B. equi* gene encoding a protein with 52% amino acid identity and 68% amino acid similarity with EMA-1 were obtained. The deduced amino acid sequences of EMA-1 and EMA-2 are shown in Fig. 1. With the exception of the amino acid at position 34 (T) (Fig. 1), EMA-2 possesses the previously reported N-terminal sequence (DEAP*xxxx*AVVD) of a 30-kDa *B. equi* erythrocyte-stage protein [3]. The missing amino acids (indicated by *x*) (KVSG) were obtained from the deduced amino acid sequences of *ema2* clones. The cDNA encoding EMA-2 was partial, missing 14 amino acids from the N-terminus. The remaining amino acids were deduced from a genomic clone of *ema2.* The amino acid identities between EMA-1 and EMA-2 are shown in black (FIG. 1). Arrows indicate the signal peptidase cleavage site and the potential N-linked glycosylation site. As was previously shown for the *ema1* gene [3], the ema2 gene is present as a single copy (data not shown). Also, comparison of the cDNA and genomic copies of *ema2* revealed the absence of introns.

Table 1
Nucleotide and amino acid changes among *ema1* genes [GenBank accessions U97167 (Brazil) and U97168 (Morocco)]

Nucleotide	81	300	309	417	580	609	775
Brazil	T	T	C	G	G	T	T
Florida	C	T	C	G	A	C	C
Morocco	C	G	T	A	G	C	T
Amino acid change	None	None	None	None	S to G	None	L to F

3.2. Amino acid homology among T. parva, T. annulata, T. mutans, T. sergenti, T. buffeli and B. equi 30/34-kDa erythrocyte-stage proteins.

Fig. 2 shows the amino acid identity shared among *Theileria* species and *B. equi* erythrocyte-stage, immunodominant proteins. A total of 12 sequences of *T.* species proteins were compared to EMA-1 and EMA-2 of *B. equi*. Fig. 2 was constructed using a representative from each of the *T.* species. The amino acid identities shown in black were determined by applying the following definitions: (i) the amino acid must be contained in either EMA-1 or EMA-2; and (ii) the amino acid must also be in the majority (≥ 4 of 7) of the combined *B. equi* and *T.* species proteins. The mean amino acid identity among the *Theileria* proteins and either of the *B. equi* proteins is 31%. As shown in Fig. 2, the shared amino acids are distributed throughout the proteins, and there are no apparent peptide repeats. Excluding the signal peptide, with the exception of *T. annulata* and EMA-2 these proteins contain no cysteines (Fig. 2). Excluding the signal peptide, *T. annulata* and EMA-2 have a single cysteine.

Charge analysis of the amino acids comprising these proteins revealed the following isoelectric points: *T. mutans* (9.47), *T. annulata* (9.74), *T. parva* (9.43), *T. buffeli* (9.47), *T. sergenti* (9.45), *B. equi*-EMA-1 (5.52) and *B. equi*-EMA-2 (6.05). The neutral to acidic charge profile of EMA-1 and EMA-2 was unexpected considering their surface membrane location and the working hypothesis that they interact with the equine erythrocyte surface.

3.3. Genetic diversity of the B. equi ema1 gene

To analyze the degree of diversity among *ema1* genes, PCR copies of *ema1* from infected horses in Morocco and Brazil were obtained and sequenced. Comparisons of the *ema1* genes from Morocco and Brazil with *ema1*-Florida [3] revealed greater than 99% nucleotide identity. The nucleotide changes result in two amino acid changes (Table 1). These amino acid changes are the substitution of a glycine for a serine in the Morocco and Brazil EMA-1 (position 194) and a phenylalanine for a leucine in the Morocco and Brazil EMA-1 (position 259).

The potential N-linked glycosylation site (Asn-Ser-Thr) (at position number 239) is conserved among all three EMA-1 proteins. Although the Morocco and Brazil *ema1* genes share 100% amino acid identity, the genes have five nucleotide differences (Table 1).

3.4. In vitro labeling of B. equi proteins

To investigate whether EMA-1 and EMA-2 possess a glycosylphosphatidylinositol (GPI) anchor, and utilize the single predicted N-linked glycosylation site, in vitro cultures of *B. equi* were incubated separately with the following precursors: [^{3}H]palmitic acid, [^{3}H]myristic acid, [^{3}H]inositol, [^{14}C]ethanolamine, [^{3}H]glucosamine, [^{3}H]fucose, and [^{3}H]galactose. An immunoprecipitation assay using post-infection serum and mAb 36/133.97 [14] shows the incorporation of [^{3}H]palmitic acid into both EMA-1 and EMA-2 (Fig. 3). Also, EMA-1 and EMA-2 incorporated [^{3}H]myristic acid, [^{3}H]inositol, and [^{3}H]glucosa-

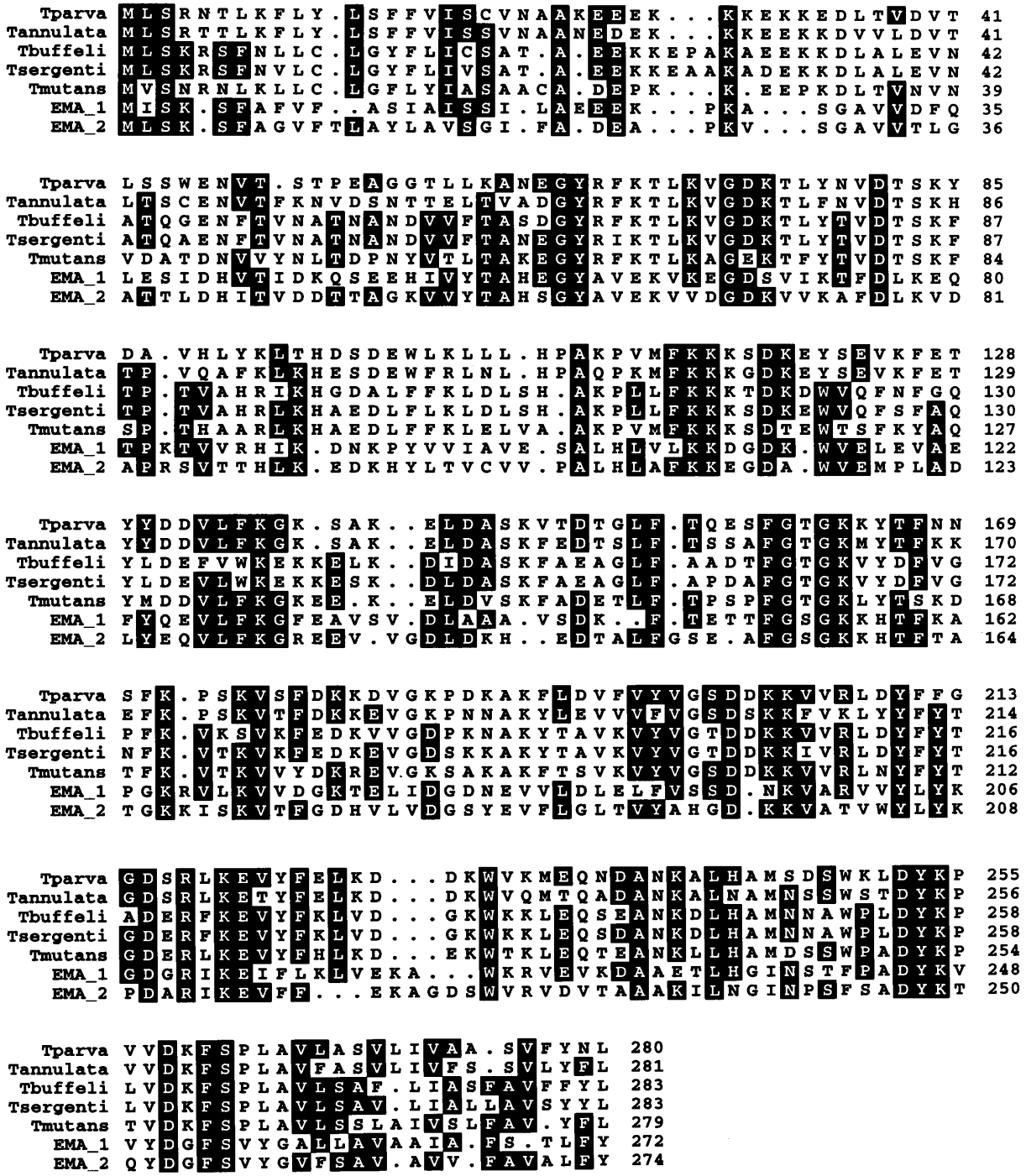

Fig. 2. Alignment of the deduced amino acid sequences of *B. equi* erythrocyte-stage proteins EMA-1 and EMA-2 with that of a *Theileria parva* (GenBank™ accession number L47209), *T. annulata* (GenBank™ accession number S54804), *T. buffeli* (GenBank™ accession number S27859), *T. sergenti* (GenBank™ accession number S27863), and *T. mutans* (GenBank™ accession number U85039) erythrocyte-stage 30–34-kDa proteins. The amino acid identities, shown in black, were determined by applying the following definitions: (1) the amino acid must be within in either EMA-1 or EMA-2; and (2) the amino acid must also be in the majority of the combined *B. equi* and *T.* species proteins. Gaps are represented by dots.

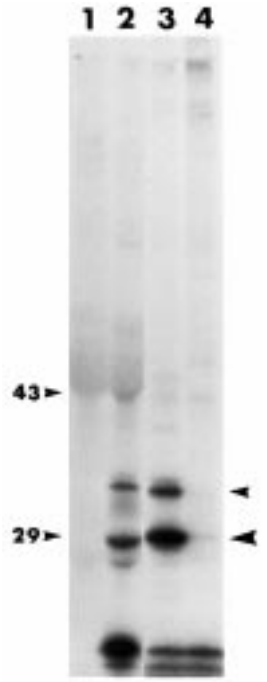

Fig. 3. Immunoprecipitation of *B. equi* proteins labeled in vitro with [^{3}H]palmitic acid. Lanes 1–4 represent precipitates by 10 μl of undiluted preinfection equine serum; 10 μl of undiluted postinfection equine serum; 10 μg ml^{-1} of mAb 36/133.97 [14] and 10 μg ml^{-1} of isotype control mAb, respectively. EMA-1 and EMA-2 are indicated by arrows (small and large, respectively).

mine (data not shown). However, EMA-1 and EMA-2 apparently did not incorporate [^{14}H]ethanolamine, [^{3}H]fucose or [^{3}H]galactose (data not shown). Fig. 3 shows that EMA-1 and EMA-2 are [^{3}H]palmitic acid labeled proteins immunoprecipitated by post-infection equine serum. Additional [^{3}H]palmitic acid labeled proteins recognized are 97, 27, 18 and 17-kDa (Fig. 3). Post-infection serum also bound labeled proteins of 97 and 43 kda (myristic acid) and 18 kDa (inositol) (data not shown). However, regardless of the label incorporated, equine immune sera bound consistently to EMA-1 and EMA-2.

Due to the fact that [^{3}H]glucosamine will incorporate into GPI anchors, [^{3}H]fucose and [^{3}H]galactose were tested for incorporation into EMA-1 and EMA-2. Although EMA-1 and EMA-2 incorporated [^{3}H]glucosamine, incorporation of [^{3}H]fucose or [^{3}H]galactose was not detected in any proteins. Although EMA-1 and EMA-2 contain a single predicted N-linked glycosylation site at amino acids 241–244, EMA-2 contains a prohibited amino acid (proline) at position 242 in the consensus sequence Asn-Xaa-Ser/Thr/Cys [15].

3.5. Relationship of EMA-1, EMA-2 and T. species erythrocyte-stage 30/34-kDa proteins to Elapid hemolysins

A profile search [16] for functional homology with the *B. equi* and *T.* species proteins identified a region of homology with Elapid cytotoxin proteins (Fig. 4). A subset of Elapid cytotoxins has been shown to bind to and lyse erythrocytes [17,18]. These cytotoxins are hemolytic protein 12B and 12A (cytotoxin 1 and 2) from *Hemachatus haemachatus* [18]. This region of 22 and 23% amino acid identity (EMA-1 and EMA-2, respec-

```
cx1_hemha  .LKCHNKLVPFLSKTCPEGKNLCYKMTMLKMPKIPI.KRGCTDACPKSSLL.VKVVCCNKDKCN  61
cx2_hemha  .LKCHNKVVPFLSKTCPEGKNLCYKMTLKKVPKIPI.KRGCTDACPKSSLL.VNVMCCKTDKCN  61
 Tbuffeli  ASKFAEAGL.FAADTFGTGKVYDFVGPFK.VKSVKFEDKVVGDP.KNAKYTAVKVYVGTDDKK.  207
Tsergenti  ASKFAEAGL.FAPDAFGTGKVYDFVGNFK.VTKVKFEDKEVGDS.KKAKYTAVKVYVGTDDKK.  207
   Tparva  ASKVTDTGL.FTQESFGTGKKYTFNNSFK.PSKVSFDKKDVGKP.DKAKFLDVFVYVGSDDKK.  204
Tannulata  ASKFEDTSL.FTSSAFGTGKMYTFKKEFK.PSKVTFDKKEVGKP.NNAKYLEVVVFVGSDSKK.  205
  Tmutans  VSKFADETL.FTPSPFGTGKLYTSKDTFK.VTKVVYDKKEVGKS.AKAKFTSVKVYVGSDDKK.  203
    EMA_1  LAAA.AVSDK.FTETTFGSGKKHTFKAPGKKVLKVVDGKTELIDG.DNEVVLDLELFVSSDNK..  197
    EMA_2  LDKHEDTAL.FGSEAFGSGKKHTFTATGKKISKVTFGDKVLVDG.SYEVFLGLTVYAHGDKK..  199
```

Fig. 4. Alignment of the deduced amino acid sequences of *B. equi* erythrocyte-stage proteins EMA-1 and EMA-2 with that of a *Theileria parva*, *T. annulata*, *T. buffeli*, *T. sergenti* and *T. mutans* erythrocyte-stage 30–34-kDa proteins, and Elapid hemolytic proteins 12B and 12A (cx1 and cx2)(GenBank™ accession numbers P01471 and P24776). Identicai residues are shown in black and similar residues in light grey. The GCG program PILEUP was used to align the sequences, and PRETTYBOX was used to display the alignment.

tively) is located between amino acids 140 and 200 of EMA-1 and EMA-2 (Fig. 4).

The GCG program GAP was used to determine a quality score of 17 for the amino acid homology between *Hemachatus haemachatus* cytotoxins 1 and 2 and EMA-1 and EMA-2. An average quality score of ten was derived from ten randomizations of the same area of amino acid homology. This narrow difference indicates the amino acid homology is of limited significance. Importantly, evaluation of the amino acid homology reveals that many of the amino acid identities are lysines (6/13 for EMA-1 and 6/14 for EMA-2).

Similar to the Elapid cytotoxins, the *T*. species erythrocyte proteins in this region possess isoelectric points of 9.9–10.06. While this region of *B. equi* EMA-1 and EMA-2 possess isoelectric points of 5.85 and 6.97, respectively, the isoelectric point for a narrowed peptide in this region (amino acids 150–170) of EMA-1 and EMA-2 possess isoelectric points of 11.36 and 10.97, respectively. Complete analysis of EMA-1 and EMA-2 peptide charge profiles revealed this region to be the only one within these proteins with a basic charge profile. This peptide in EMA-1 is FTETTFGSGKKHTFKAPGKRVLKV.

Analysis of the proteins using the GCG program MOTIFS did not reveal any significant scores between the *T*. species proteins or *B. equi* EMA-1 and EMA-2 and known enzyme consensus sequences.

4. Discussion

Equi merozoite antigens-1 and -2 (EMA-1 and EMA-2) are erythrocyte-stage proteins of *B. equi* which possess a surface epitope that is conserved worldwide and induce specific high antibody levels [2,3,14]. Previous data showed that EMA-1 shared greater than 30% identity with 30–34-kDa proteins from *T. sergenti* and *T. buffeli* [3]. Results of the present study show that EMA-1 shares 52% amino acid identity with a second 30-kDa *B. equi* erythrocyte-stage protein, EMA-2, and 31% amino acid identity with 30–34-kDa proteins of *T. parva*, *T. annulata* and *T. mutans*. Additionally, EMA-1 and EMA-2 and the 30–34-kDa *T*. species proteins share a region of approximately 22% amino acid identity with two Elapid cytotoxin proteins. These cytotoxins, which bind to and lyse erythrocytes, are from *Hemachatus haemachatus* [17,18]. Although the *B. equi* and *T*. species 30–34-kDa proteins share genetic and structural similarities, they diverge significantly in their overall charge profiles.

Consistent with their location on the surface of the erythrocyte-stage parasite, EMA-1 and EMA-2 possess a GPI anchor [19,20]. This conclusion is based on immunoprecipitation of EMA-1 and EMA-2 labeled with [^{3}H]palmitic acid, [^{3}H]myristic acid, and [^{3}H]inositol with postinfection equine serum. Previous work had shown that a protein epitope shared by EMA-1 and EMA-2 was immunodominant for antibody induction and shared by isolates worldwide [14].

The use of the single putative N-linked glycosylation site in EMA-1 and EMA-2 was not proven. Although EMA-1 and EMA-2 incorporated glucosamine, incorporation of fucose or galactose was not detected. Glucosamine can also be incorporated into GPI anchors [21]. The use of the putative N-linked glycosylation site of EMA-2 is unlikely since there is a prohibited amino acid (proline) at position 242 in the consensus sequence Asn-Xaa-Ser/Thr/Cys [15]. The predicted molecular mass of native EMA-2 is approximately 32 kDa (equivalent to the molecular mass of the deduced amino acids (34.8 kDa) minus the signal peptide (2.7 kDa) and the approximate hydrophobic tail (3.1 kDa), with 3 kDa added for the GPI anchor [22]). The predicted molecular mass for EMA-2 of 32 kDa is in excess of the apparent molecular mass of native EMA-2 (29 kDa) in SDS-PAGE (Fig. 3).

Comparison of the apparent molecular mass of EMA-1 from SDS-PAGE (34 kDa) (Fig. 3) with the predicted molecular mass from the deduced amino acid sequence (32 kDa) indicates the N-linked glycosylation site at position 241–244 may be utilized. N-linked glycosylation would add approximately 2.1 kDa per used site [23]. Alternatively, the numerous serines and

threonines in EMA-1 may serve as targets for O-linked glycosylation similarly to that reported for *Plasmodium falciparum* [24]. Although EMA-1 and EMA-2 clearly possess GPI anchors, the use of the putative N-linked glycosylation site within these proteins remains unresolved. Other work questions whether N-linked glycosylation occurs at all in other members of the phylum Apicomplexa [24].

Regardless of the use of the predicted N-linked glycosylation site in EMA-1, the conservation of this single site is in contrast to that reported for the *T*. species 30–34-kDa proteins [4]. Previous analysis of N-linked glycosylation sites of 30–34-kDa proteins of *T. annulata*, *T. sergenti*, *T. buffeli* and *T. parva* suggested marked diversity in potential N-linked glycosylation sites [4]. Analysis of three *ema1* genes from Florida, Morroco, and Brazil showed that the putative N-linked glycosylation site at position 241–244 was conserved as the only site. The conservation of the potential N-linked glycosylation site of EMA-1 is in agreement with the 99% amino acid conservation of EMA-1 among the Florida, Morroco and Brazil *ema1* genes. Although evaluation of three geographically distinct copies of the *ema1* gene showed limited variation, the data can not be extended to conclude that all *ema1* gene copies are conserved at the 99% level. The *ema1* genes evaluated were obtained from clinically normal, persistently infected animals. In order to strengthen this observation, copies of the *ema1* gene from parasite populations of additional carrier horses and from horses during acute disease in clinically affected animals should also be evaluated.

The precise relationship of the *ema1* and *ema2* genes is not defined by the data presented here. High stringency hybridizations of restriction digests of *ema1* [3] and *ema2* indicate these are single copy genes. Since the data presented was derived from a mixed parasite population and not a biologically cloned parasite, we do not know if each haploid erythrocyte-stage parasite contains both genes or if sub-populations possess only *ema1* or *ema2*. If each haploid parasite contains *ema1* and *ema2*, it is probably appropriate to define them as separate, but related genes. However, if *ema1* and *ema2* exist separately in parasite sub-populations they should be referred to as alleles.

Structurally, the *B. equi* and *T*. species 30–34-kDa proteins possess a signal peptide [3,4], lack apparent antigenic repeats, and excluding the signal peptide, contain one or no cysteines. However, charge analysis of these 30–34-kDa antigens showed that the *T*. species proteins possess isoelectric points >9 while, unexpectedly, the *B. equi* proteins possess acidic isoelectric points. The acidic charge profile of EMA-1 and EMA-2 was unexpected considering their surface location and our working hypothesis that these proteins interact with the equine erythrocyte surface. The acidic character of EMA-1 and EMA-2, and the finding of amino acid homology with Elapid cytotoxin proteins directed our analysis to discrete peptide regions of EMA1 and EMA-2.

A subset of Elapid cytotoxins has been shown to bind and lyse erythrocytes [17,18]. These cytotoxins are hemolytic proteins 12B and 12A (cytotoxin 1 and 2) from *Hemachatus haemachatus* [18]. A region of approximately 22% amino acid identity is shared with a similar region of the 30–34-kDa *T*. species proteins and EMA-1 and EMA-2. The region of amino acid identity is located between amino acids 140 and 200 of EMA-1 and EMA-2. Analysis of this region shows that the majority of the amino acid identity is attributed to the basic amino acid lysine (Fig. 4). Isoelectric point analysis of this region in EMA-1 and EMA-2 revealed an acidic charge. However, when the region analyzed for charge was limited to amino acids 150–170 of EMA-1 and EMA-2, isoelectric points of 11.36 and 10.97 were calculated for these peptides of EMA-1 and EMA-2, respectively. Further analysis revealed that within EMA-1 and EMA-2, this region of amino acids, 150–170, is the only linear peptide with an overall basic charge. This has led us to hypothesize that this region may be important in interacting with the erythrocyte surface.

The collective data presented shows that EMA-1 and EMA-2 are GPI anchored proteins consistently bound by post-infection serum.

Also, these *B. equi* proteins share significant amino acid homology with a similar family of erythrocyte-stage 30–34-kDa proteins of *T.* species. In addition to the shared genetic homology, structurally these *B. equi* and *T.* species proteins are similar in that they do not contain antigenic repeats and, excluding the signal peptide, contain one or no cysteines. The conservation of surface proteins across species of parasites suggests important functional attributes, and combined with immunodominance for antibody, merits intense evaluation of their role in stimulating protective immune responses.

Acknowledgements

The authors thank T. Harkins for nucleotide sequencing, and Nancy Benson for assistance in the screening of the cDNA library for the *ema2* gene. We also thank Dr Luiz Shozo Ozaki of the Center of Biotechnology, Universidade Federal do Rio Grande do Sul for *B. equi* DNA from Brazil and Drs Hamid Sahbi and Abdelkebir Rhalem of the Department de Parasitologie, Institut Agronomique et Vétérinaire Hassan II, Rabat for *B. equi* DNA from Morocco. This work was supported by the US Department of Agriculture, Animal Plant Health Inspection Service Cooperative (CWU 5348-34000-004-01) and the US Department of Agriculture, Agricultural Research Service (CWU 5348-34000-004-00D).

References

[1] Knowles DP Jr., Kappmeyer LS, Perryman LE. Specific immune responses are required to control parasitemia in *Babesia equi* infection. Infect Immun 1994;62:1909–13.

[2] Knowles DP Jr. Control of *Babesia equi* parasitemia. Parasitol Today 1996;12:195–8.

[3] Kappmeyer LS, Perryman LE, Knowles DP. A *Babesia equi* gene encodes a surface protein with homology to *Theileria* species. Mol Biochem Parasitol 1993;62:121–4.

[4] Shiels BR, d'Oliveira C, McKellar S, Ben Miled L, Kawazu S, Hide G. Selection of diversity at putative glycosylation sites in the immunodominant merozoite/piroplasm surface antigen of *Theileria* parasites. Mol Biochem Parasitol 1995;72:149–62.

[5] D'Oliveira C, Tijhaar EJ, Shiels BR, van der Weide M, Jongejan F. Expression of genes encoding two major *Theileria annulata* merozoite surface antigens in *Escherichia coli* and a *Salmonella typhimurium* aro A vaccine strain. Gene 1996;12:33–9.

[6] Kawazu S, Sugimoto C, Kamio T, Fujisaki K. Analysis of the genes encoding immunodominant prioplasm surface proteins of *Theileria sergenti* and *Theileria buffeli* by nucleotide sequencing and polymerase chain reaction. Mol Biochem Parasitol 1992;56:169–76.

[7] Matsuba T, Sugimoto C, Hattori M, Sako Y, Fujisaki K, Onuma M. Expression of a 32 kilodalton *Theileria sergenti* piroplasm surface protein by recombinant baculoviruses. Int J Parasitol 1995;25:939–43.

[8] Kubota S, Sugimoto C, Onuma M. Population dynamics of *Theileria sergenti* in persistently infected cattle and vector ticks analysed by a polymerase chain reaction. Parasitology 1996;112:437–42.

[9] Matsuba T, Kubota H, Tanaka M, Hattori M, Murata M, Sugimoto C, Onuma M. Analysis of mixed parasite populations of *Theileria sergenti* using cDNA probes encoding a major piroplasm surface protein. Parasitology 1993;107:369–77.

[10] Dickson J, Shiels BR. Antigenic diversity of a major merozoite surface molecule in *Theileria annulata*. Mol Biochem Parasitol 1993;57:55–64.

[11] Knowles DP Jr., Kappmeyer LS, Stiller D, Hennager SG, Perryman LE. Antibody to a recombinant merozoite protein epitope identifies horses infected with *Babesia equi*. J Clin Microbiol 1992;30:31223126.

[12] Gribskov M, McLachlan AD, Eisenberg D. Profile analysis: detection of distantly related proteins. Proc Natl Acad Sci USA 1987;84:43554358.

[13] Holman PJ, Chieves L, Frerichs WM, Olson D, Wagner GG. *Babesia equi* erythrocytic state continuously cultured in an enriched medium. J Parasitol 1994;80:232–6.

[14] Knowles DP Jr., Perryman LE, Goff WL, Miller CD, Harrington RD, Gorham JR. A monoclonal antibody defines a geographically conserved surface protein epitope of *Babesia equi* merozoites. Infect Immun 1991;59:2412–7.

[15] Gooley AA, Williams KL. How to find, identify and quantitate the sugars on proteins. Nature 1997;385:557–9.

[16] Altschul SF, Gish W, Miller W, Meyers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403–10.

[17] Dufton MJ, Hider RC. The structure and pharmacology of elapid cytotoxins. In: Harvey AL, editor. Snake Toxins. New York: Pergamon, 1991:259-302.

[18] Fryklund L, Eaker D. Complete amino acid sequence of a nonneurotoxic hemolytic protein from the venom of *Haemachatus haemachatus* (African ringhals cobra). Biochemistry 1973;4:661–7.

[19] McConville MJ, Ferguson MAJ. The structure, biosynthesis and function of glycosylated phosphatidylinositols

in the parasitic protozoa and higher eukaryotes. Biochem J 1993;294:305–24.
[20] Cross GAM. Glycolipid anchoring of plasma membrane proteins. Annu Rev Cell Biol 1990;6:1–39.
[21] Varki A. Metabolic radiolabeling of glycoconjugates. In: Lennarz WJ, Hart GW, editors. Methods in Enzymology, vol. 230, Guide to Techniques in Bycobiology. San Diego: Academic Press, 1994:16–32.
[22] Schofield L, Hackett F. Signal transduction in host cells by a glycosylphosphatidylinositol toxin of malaria parasites. J Exp Med 1993;177:145153.
[23] Lennarz WJ. Lipid-lined sugars in glycoprotein synthesis. Science 1975;188:986–91.
[24] Dieckmann-Schuppert A, Bender S, Odenthal-Schnittler M, Bause E, Schwarz RT. Apparent lack of N-glycosylation in the asexual intraerythrocytic stage of *Plasmodium falciparum*. Eur J Biochem 1992;205:815 825.